Merrill Rudd

FSH 507 Spatial-temporal models

Homework 1

In this exercise, I compared three different generalized linear models to predict catch rates for Eastern Bering Sea pollock. The first model assumed that the catch arises from a delta-lognormal distribution, estimating an intercept parameter, a shape parameter associated with the probability of catching zero pollock, and the log of the standard deviation for the lognormal distribution. The second model is a delta-gamma distribution, estimating an intercept parameter, parameters associated with covariates on longitude and latitude, a shape parameter associated with the probability of catching zero pollock, and the log of a standard deviation related to the gamma shape and scale parameters. The model fits to the Eastern Bering Sea pollock catch data are shown in Figure 1.

The delta-gamma model that included latitude and longitude covariates to predict catch had the lowest negative log-likelihood and the highest predictive probability according to K-fold cross-validation (Table 1). In this case, using the information contained in the latitude and longitude does an overall better job at predicting catch rates than the models that did not consider these covariates.

The simulation test confirmed that the estimation models have the best performance (i.e. are able to estimate the parameters most accurately) when the data arise from a matching process (Figure 2). While the delta-gamma estimation model has the best predictive ability with real data, it has the lowest precision when used as an estimation model across generated datasets simulated with various simulation models. An interesting point, however, is that the delta-gamma model with covariates is still unbiased when the data is generated from a delta-gamma model that does not consider latitude and longitude. While the estimates are biased when the data truly arise from a delta-lognormal function, the low precision means that the distribution of estimates out of 100 includes the true value of the intercept parameter. Overall, the delta-lognormal and delta-gamma estimation models have approximately the same level of precision, and higher precision than the delta-gamma with covariates model. However, these models tend to be more biased than the delta-gamma with covariates model when the simulation models and estimation models do not match.

Table 1. Comparison of the negative log-likelihood, number of parameters, and cross-validation score for the three models explored in this exercise.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | NLL | # Parameters | Cross-validation score |
| delta-lognormal | 62639 | 3 | 5.172 |
| delta-gamma | 61922 | 3 | 5.198 |
| delta-gamma with lat/lon covariates | 60713 | 5 | 4.672 |

Figure 1. Model fits to catch data for the Eastern Bering Sea pollock. Observed catch data is shaded in pink, model predictions for the delta-lognormal are blue, delta-gamma green, and delta-gamma with covariates (covar) are yellow. The model predictions were generated by simulating 12,210 predictions using a simulation model matching the estimation model and the estimated parameters for each respective model.

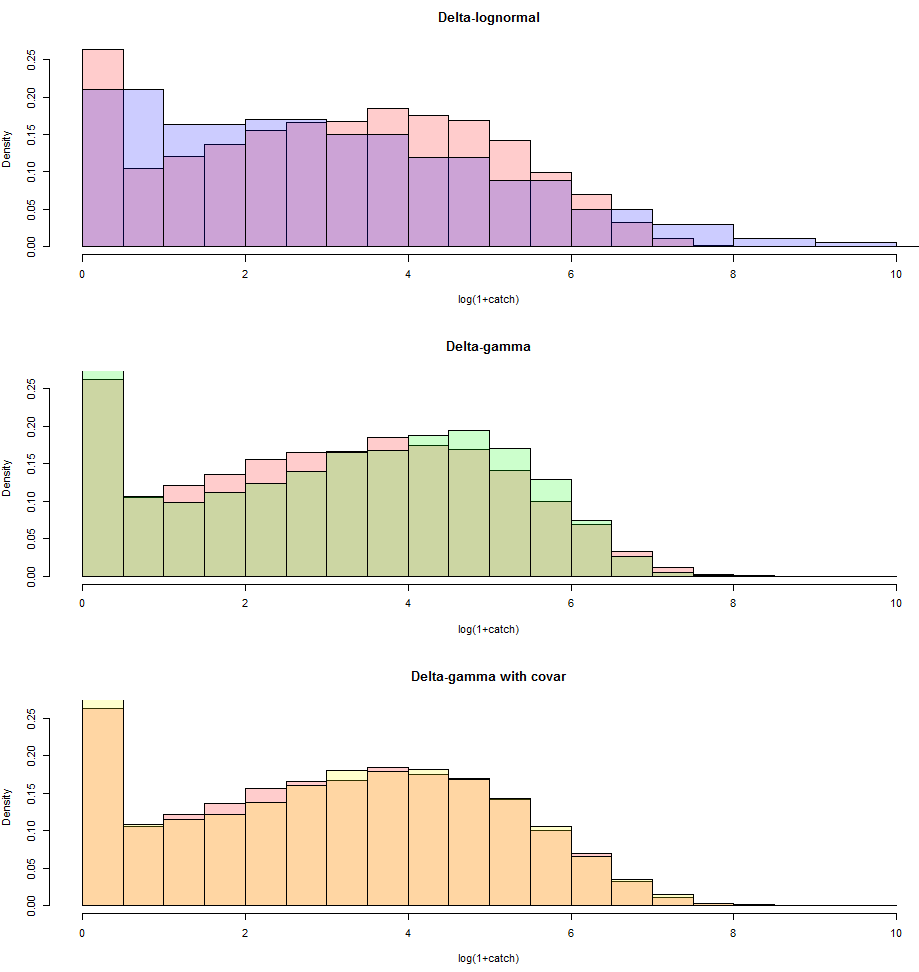


Figure 2. Distributions of estimates of the intercept parameter from 100 iterations of simulated data for each of the three simulation models, from each of the three estimation models. The dotted lines show where the estimates would be unbiased compared to the true value. The red distributions were generated from the delta-lognormal model, light blue distributions from the delta-gamma model, and dark blue distributions from the delta-gamma model with covariates for latitude and longitude. The dark black line shows the median of each distribution. The colored shading corresponds to the estimation model that was used on each simulated dataset. The best accuracy occurs when the simulation model matches the operating model.

